Forecasting the Equity Premium

Bernt Arne Ødegaard

24 November 2021

Contents

1 The Equity Market Premium **1**

2 Is the equity market premium predictable? **1**
2.1 How predictable can the market be? 1
2.2 Empirically investigating predictability 2

3 Replicating Goyal Welch **2**

4 Replicating Cooper Priestley **6**
4.1 R codes 10
4.1.1 Reading the data 10
4.1.2 Doing the analysis 10

5 Literature **13**

1 The Equity Market Premium

Is the difference between the return on the stock market and a risk free interest rate

$$r_{m,t} - r_{f,t}$$

In practice we use the return of a broad based stock market index to proxy for the stock market, and a treasury rate (often long term) to proxy for the risk free rate.

Most estimates of the equity market risk premium puts in the range of 5-7%. This is viewed as high, and lead to the "equity premium puzzle."

Much of the empirical literature is concerned with ways of estimating the market risk premium, which is hard (Merton, 1980).

In this lecture we are concerned with empirical method for *predicting* the equity market premium.

2 Is the equity market premium predictable?

There is a large literature on the predictability of (US) market indices, essentially asking: Can we predict the equity market premium? The literature is macro-finance in tone, trying to predict aggregate stock market indices, not individual stocks.

2.1 How predictable can the market be?

Starting point: Classical efficient market tests: Martingale hypothesis.

Seem to indicate no predictability

However, if we allow for time varying risk preferences, can have some return predictability, corresponding to changes in risk premia.

Can find an upper bound, starting from first principles

$$p_t = E[m_{t+1}P_{t+1}]$$

2.2 Empirically investigating predictability

Cute picture showing the time series of this research, from Henkel, Martin, and Nardari (2011):

562 S.J. Henkel et al. / Journal of Financial Economics 99 (2011) 560–580

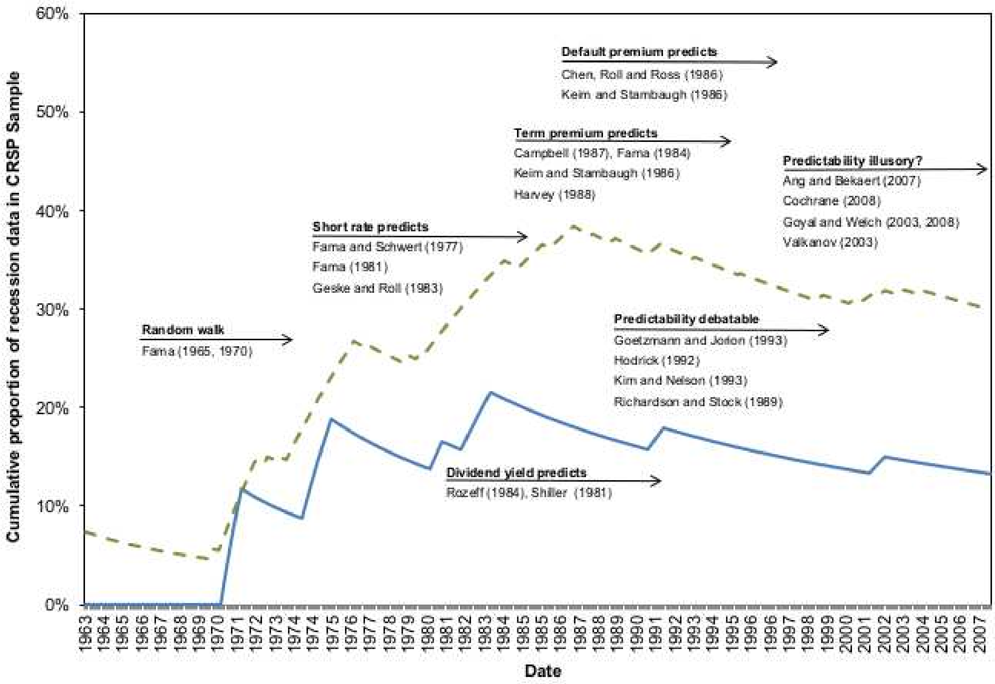


Fig. 2. The time-series of predictability research. The literature on stock return predictability follows closely the availability of recession data as a cumulative proportion of the total data in CRSP which originally started in 1962. Shown are the percentages of recession data as a percentage of the available data at a given date, as measured by NBER (solid line) and RSVAR (dashed line) dates. Both the NBER and RSVAR samples show similar profiles, although RSVAR recession probabilities represent a much larger proportion of the data. Many seminal, and first, papers on return predictability were published just after the peaking of the proportion of recession data to total available data in 1985 and are followed by a decline in the proportion of recession data thereafter. The citations are representative for expository purposes and are not intended to be indicative of initial research, nor a comprehensive literature survey (Ang and Bekaert, 2007; Campbell, 1987; Chen et al., 1986; Cochrane, 2008; Fama, 1965, 1970, 1981, 1984; Fama and Schwert, 1977; Geske and Roll, 1983; Goetzmann and Jorion, 1993; Goyal and Welch, 2003; Harvey, 1988; Hodrick, 1992; Keim and Stambaugh, 1986; Kim and Nelson, 1993; Richardson and Stock, 1989; Rozeff, 1984; Shiller, 1981; Valkanov, 2003; Welch and Goyal, 2008).

Also shows how a single picture can be used to give an interesting argument, with the lines on fraction of data with recessions, together with the arguments that the predictability is coming from recession periods.

Since the data is available for the Goyal and Welch (2008) piece, can use that data to replicate their results, to understand the methods used in this type of analysis.

3 Replicating Goyal Welch

To get used to working with these kinds of issues, we will replicate (some of) the analysis of Goyal and Welch (2008), primarily because their data is readily available, and we can compare our work with their numbers and figures.

We will use their annual data. It is available from the RFS web cite, or from Amit Goyal's web page. There is also a set of updated data series.

The following reads the data into zoo series

```
library(zoo)
dataGoyalWelchAnnual <- read.table("../data/PredictorData_annually.csv",
                                    header=TRUE,sep=",",na.strings=c("NaN"))

dataGoyalWelchAnnual <- zoo(dataGoyalWelchAnnual[,2:ncol(dataGoyalWelchAnnual)],
                            order.by=dataGoyalWelchAnnual[,1])
```

This is an overview of the data.

```
> head(dataGoyalWelchAnnual)
     Index  D12  E12 b.m tbl AAA BAA lty cay ntis      Rfree infl eqis ltr
1871  4.74 0.26 0.40  NA  NA  NA  NA  NA  NA   NA 0.05733672   NA   NA  NA
1872  5.07 0.30 0.43  NA  NA  NA  NA  NA  NA   NA 0.07189866   NA   NA  NA
1873  4.42 0.33 0.46  NA  NA  NA  NA  NA  NA   NA 0.08661904   NA   NA  NA
1874  4.54 0.33 0.46  NA  NA  NA  NA  NA  NA   NA 0.04854423   NA   NA  NA
1875  4.37 0.30 0.36  NA  NA  NA  NA  NA  NA   NA 0.04380927   NA   NA  NA
1876  3.58 0.30 0.28  NA  NA  NA  NA  NA  NA   NA 0.04106633   NA   NA  NA
     corpr svar csp ik CRSP_SPvw CRSP_SPvwx
1871    NA   NA  NA NA        NA         NA
1872    NA   NA  NA NA        NA         NA
1873    NA   NA  NA NA        NA         NA
1874    NA   NA  NA NA        NA         NA
1875    NA   NA  NA NA        NA         NA
1876    NA   NA  NA NA        NA         NA
```

It is not obvious from this exactly what series is what, in particular it is not obvious that Index is an index of stock market prices, but here are the necessary calculations for doing the first few figures, those involving dividends and earnings.

```
Sp <- dataGoyalWelchAnnual$Index
D12 <- dataGoyalWelchAnnual$D12
E12 <- dataGoyalWelchAnnual$E12
rf  <- dataGoyalWelchAnnual$Rfree
Rf  <- 1+rf
Rm <- log(Sp+D12)-lag(log(Sp),-1)

eRm <- na.omit(Rm-log(Rf))
dp  <- na.omit(log(D12)-log(Sp))
dy  <- na.omit(log(D12)-lag(log(Sp),-1))
ep  <- na.omit(log(E12)-log(Sp))

names(eRm) <- "eRm"
names(dp)  <- "dp"
names(dy)  <- "dy"
names(ep)  <- "ep"
```

Now, most of the discussion in Goyal and Welch is concerned with comparing the outcome of a prediction exercise

$$er_{m,t} = a + bpred_{t-1} + e_t$$

where $pred_{t-1}$ is some variable thought to predict the equity market premium, to a "naive" forecast using just the historical mean of the equity market premium.

One of the metrics they use is to compare the difference in aggregate prediction errors

Goyal and Welch do both

an "in sample" analysis, asking how one would have done if one had the whole history

and

an "out of sample" analysis, asking which would have done better in predicting the equity premium, if only using data

Let us look at the code for producing the difference in sample

```
library(dyn)

in_sample_diff_calc_annual <- function(predictor){
                                    # for annual data start when having 21 years of data
  first_obs <- 21
  data <- merge(eRm,predictor,all=FALSE)
  names(data) <- c("eRm","predictor")
```

```
    demeaned2 <- zoo((as.numeric(data$eRm - mean(data$eRm))^2),
                     order.by=index(data$eRm))

    regr <- dyn$lm(data$eRm ~lag(data$predictor,-1))
    res2 <- as.numeric(regr$residuals^2)
    res2 <- zoo(res2,order.by=index(data$eRm)[-1])
                                          # there is one less residual than mean differences
    tmp <- merge(demeaned2,res2,all=FALSE)
    diff <- cumsum(tmp$demeaned2)-cumsum(tmp$res2)
    diff <- diff - as.numeric(diff[first_obs])
                                          # to align at zero on the first oos observation
    return (diff)
}
```

And then out of sample

```
out_of_sample_diff_calc_annual <- function (pred) {
                                          # for annual data start when having 21 years of data
    first_obs <- 21
    data <- merge(eRm,pred,all=FALSE)
    names(data)<-c("eRm","pred")
    head(data)
    se_NULL <- NULL
    se_ALT <- NULL
    n <- length(data$eRm)
    for (t in first_obs:n){
        se0 <- (data$eRm[t]-mean(data$eRm[1:(t-1)]))^2
        se_NULL <- rbind(se_NULL,zoo(se0,index(data$eRm)[t]))

        prem <- data$eRm[2:(t-1)]
        pred <- data$pred[1:(t-2)]
        regr <- lm(prem~pred)
        npred <- data.frame(pred <- data$pred[t-1])
        pr <- predict.lm(regr, npred)
        se1 <-( data$eRm[t] - pr)^2
        se_ALT <- rbind(se_ALT,zoo(se1,index(data$eRm)[t]))
    }
    cse_diff <- cumsum(se_NULL)-cumsum(se_ALT)
    return (cse_diff)
}
```

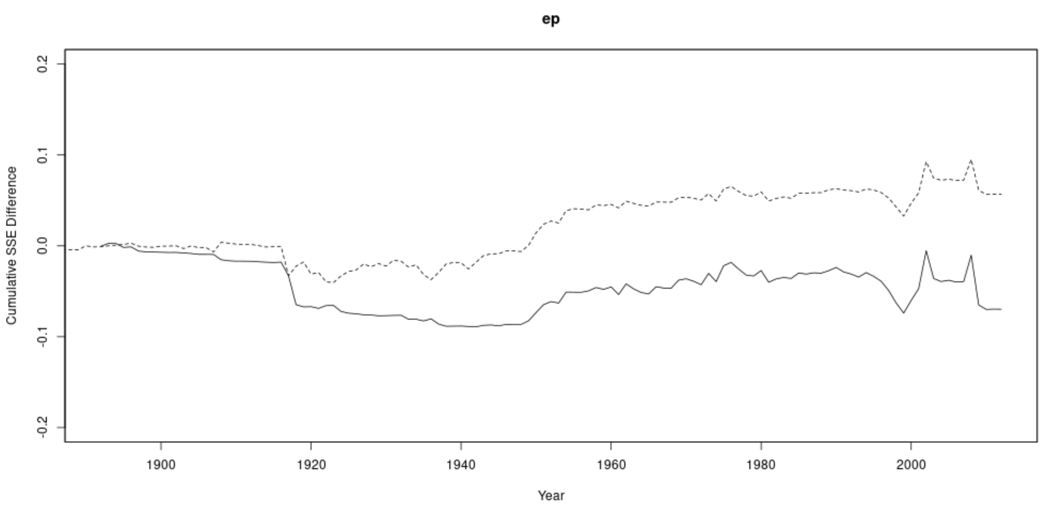
Let us now show the case of doing the calculation for dp as a predictor

```
is_diff<- in_sample_diff_calc(dp)
oos_diff<- oos_calculation(dp)

postscript("../R_plots/annual_prediction_performance_dp.eps",horizontal=FALSE,width=10,height=5)
plot(oos_diff,ylim=c(-0.2,0.2),main="dp",xlab="Year",ylab="Cumulative SSE Difference",type="l")
lines(is_diff,ylim=c(-0.2,0.2),type="l",lty=2)
dev.off()
```

Produces the results:

Doing the same for dy and ep:



4 Replicating Cooper Priestley

To illustrate that constructing the similar pictures to Goyal and Welch can add to understanding, let us look at a similar article, and show how the pictures add to our understanding.

In Cooper and Priestley (2008) it is shown that the *output gap*, a measure of the difference between the *capacity* for output relative to *actual* output.

Let us try to replicate (and extend) their results.

They construct several different measures of output gap.

The first two uses monthly date on industrial producion, and measure the output gaps as deviations from trends.

The first is the residual in the following quadratic trend regression

$$y_t = a + b \cdot t + c \cdot t^2 + v_t$$

Here y_t is the log of the industrial production index.

The second is the residual in the following trend regression

$$y_t = a + b \cdot t + c \cdot t^2 + v_t$$

The third uses data on GDP, and subtracts the actual ex post GDP from an estimated of the *potential* GDP for the US estimated by the Congressional Budget Office.

We download data from FRED, the data service of the St. Louis Federal Reserve. The industrial production (INDPR) is a monthly index. The GDP (GDP) and the Potential GDP (NGDPPOT) are both quarterly series.

The three estimated output gap series are shown in figure 1. Note the the data includes data up till 2013, so it extends the Cooper and Priestley (2008) data, which ended in 2005. Observe that the 2008 crisis has had large effects on the estimates.

As an estimate of the equity risk premium we use the monthly series RMRF provided by Ken French. Quarterly premia are calculated by adding the monthly premia.

To test for predictability we calculate an *in sample* predictive regression

$$eR_m = \alpha + \beta Gap_{t-2}$$

Note that one lags the predictive variable two periods, to make sure it is observed at the time the forecast is made. The results are shown in table 1.

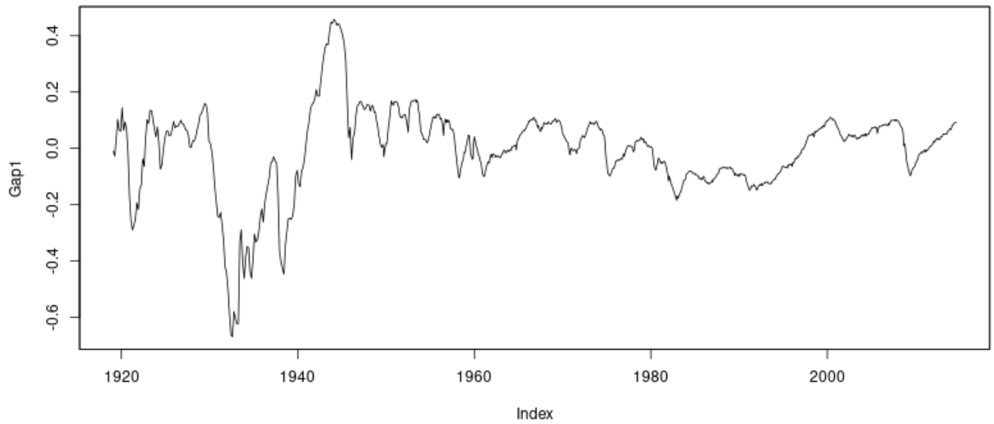
For all three estimates of Gap we find significant in-sample predictability.

To gain some understanding of what is driving the results, we use the approach of Goyal and Welch (2008), comparing the forecasting of the equity premium using this variable with a simple mean. One calculates the *cumulative squared difference* of the prediction errors, and takes the difference. Figure 2 shows the results. Note that such plots were not done in the original article.

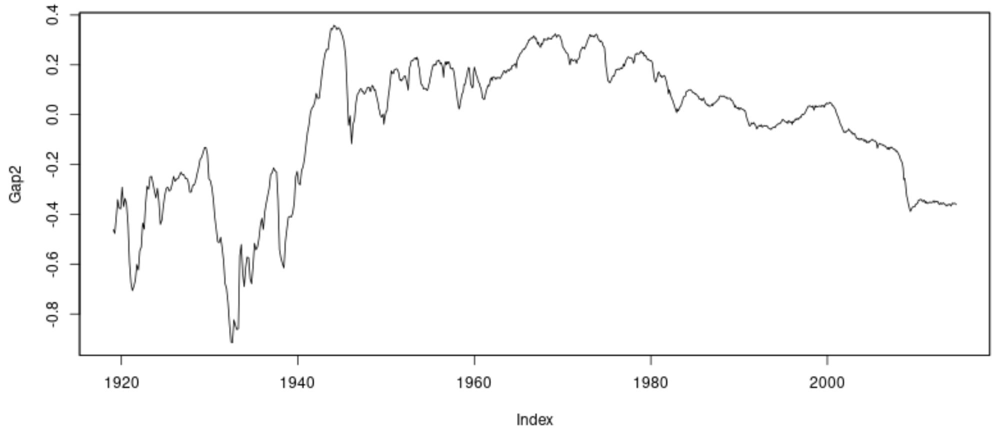
The interpretation of a figure: Ask whether the line is above zero. If it is, then the predictive regression using output gap has done better than the simple in-sample mean. All the figures end up at a positive difference, which they should, as the regressions showed predictive power. A useful extra piece of information one can get from the figures are what time periods are central in generating the predictability. Looking at the first Gap estimate, predictability is there from the very beginning. The oil crisis of '73 is a large contributor to the predictability, and the curve is flat for several decades afterward. It is only the recent crisis which is pushing the predictability upwards again. The linear trend in Gap2 is probably doing a worse job in estimating the output gap, which is behind the worse performance in panel B. The significance of the last Cap estimator, using quarterly GDP data, is very much driven by the two crises in 2000 and 2008.

Figure 1 Output Gap Series

Panel A: Gap1



Panel B: Gap2



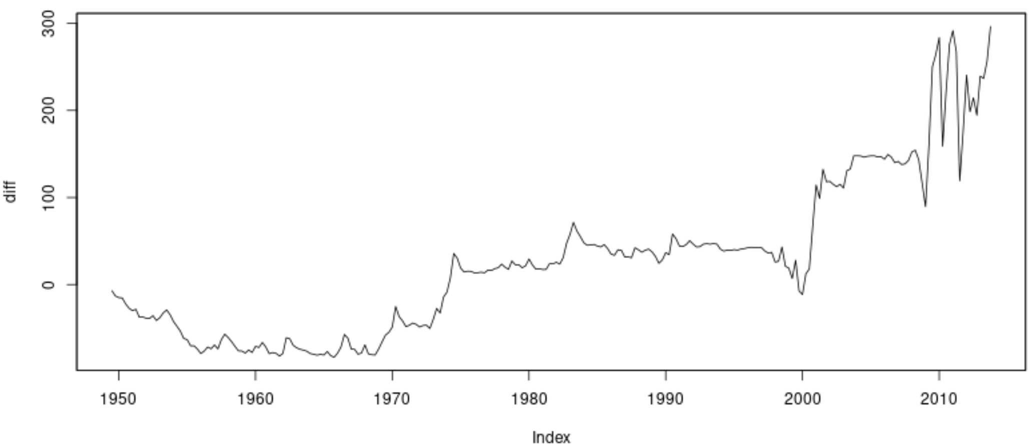
Panel C: Gap CBO

Figure 2 Predictability gain over simple mean (in sample)

Panel A: Gap1

Panel B: Gap2

Panel C: Gap CBO



Differences in squared error between resuduals of the predictive regressions, and a simple mean estimate.

Table 1 In sample regressions of predictability

	Dependent variable:		
	EqtyPrem		
	(1)	(2)	(3)
gap1	-3.664^{***}		
	(1.232)		
gap2		-1.832^{**}	
		(0.716)	
GapCBO			-0.004^{**}
			(0.002)
Constant	0.704^{***}	0.748^{***}	1.552^{***}
	(0.162)	(0.163)	(0.546)
Observations	926	926	258
Adjusted R^2	0.008	0.006	0.013
Note:			*p<0.1; **p<0.05; ***p<0.01

Results for the regression $eR_m = \alpha + \beta Gap_{t-2}$ for three different measures of output gap. Gap1 and Gap2 are calculated using monthly observations of industrial production, and are deviations from a time trend. GapCBO are calculated using quarterly observations of GDP, and is the difference between realized GDP and the estimated *potential* GDP.

4.1 R codes

4.1.1 Reading the data

```
# can replace these with direct downloads from fred, but prefer to have control over
# what data we use
library(zoo)
INDPRO <- read.table("~/data/2014/fred/INDPRO.txt",skip=40,header=TRUE)
head(INDPRO)
mIndProd <- zooreg(INDPRO$VALUE,frequency=12,start=c(1919,1))
head(mIndProd)

GDP <- read.table("~/data/2014/fred/GDP.txt",skip=19,header=TRUE)
head(GDP)
qGDP <- zooreg(GDP$VALUE,start=c(1947,1),frequency=4)
names(qGDP)<-"qGDP"
head(qGDP)

library(quantmod)
library("downloader")
TB3MS <- getSymbols("TB3MS",src="FRED")
#TB3MS <-read.table("~/data/2014/fred/TB3MS.txt", skip=11,header=TRUE)
head(TB3MS)

#NGDPPOT <- getSymbols("NGDPPOT",src="FRED")
NGDPPOT <-read.table("~/data/2014/fred/NGDPPOT.txt", skip=11,header=TRUE)
head(NGDPPOT)
qPotGDP <- zooreg(NGDPPOT$VALUE,frequency=4,start=c(1949,1))
names(qPotGDP)<-"Potential GDP"

SP500 <-read.csv("~/data/2014/yahoo_data/sp500.csv", header=TRUE)
head(SP500)

FF1 <- read.table("~/data/2014/french_data/F-F_Research_Data_Factors_monthly.txt",
                  header=TRUE,skip=3)
names(FF1)
head(FF1)
FF <- zooreg(FF1[1:4],start=c(1926,7),frequency=12)
RMRF <- FF$Mkt.RF
head(RMRF)
eRm <- RMRF
names(eRm) <- "eRm"
head(eRm)

                                    # make quarterly data, same form as others
qeRm <- aggregate(eRm,as.yearqtr,sum)
head(qeRm)
qeRm <- zooreg(coredata(qeRm),start=c(1926,3),frequency=4)
head(qeRm)
```

4.1.2 Doing the analysis

```
library(stargazer)
source("read.R")
# the output gap

lnIP <- log(mIndProd)
lnIP <- window(lnIP,start=c(1947,11))
head(lnIP)
t <- 1:length(lnIP)
t2 <- t^2
regr <- lm(lnIP ~ t + t2)

summary(regr)
Gap1 <- regr$residuals
```

```
head(Gap1)
data <- merge(eRm,lag(Gap1,-2),all=FALSE)
head(data)
EqtyPrem <- data$eRm
names(EqtyPrem)<- "eRm"
gap1 <- data[,2]
names(gap1)<- "Gap1"

regr1 <- lm(EqtyPrem ~ gap1)
summary(regr1)

demeaned_eRm<- (EqtyPrem - mean(EqtyPrem))^2
residuals <- regr1$residuals^2

head(demeaned_eRm)
head(residuals)

diff <- cumsum(demeaned_eRm)-cumsum(residuals)
#postscript(file="../../results/2014_sep_output_gap/diff_mean_Gap1.eps",horizontal=FALSE,width=10,height=5)
png(file="../../results/2014_sep_output_gap/diff_mean_Gap1.png",width=800,height=400)
plot(diff)
dev.off()


t <- 1:length(lnIP)
regr <- lm(lnIP ~ t )
summary(regr)
Gap2 <- regr$residuals
head(Gap2)
data <- merge(eRm,lag(Gap2,-2),all=FALSE)
head(data)
EqtyPrem <- data$eRm
names(EqtyPrem)<- "eRm"
gap2 <- data[,2]
names(gap2)<- "Gap2"

regr2 <- lm(EqtyPrem ~ gap2)
summary(regr2)

demeaned_eRm<-( EqtyPrem - mean(EqtyPrem))^2
residuals <- regr2$residuals^2

head(demeaned_eRm)
head(residuals)

diff <- cumsum(demeaned_eRm)-cumsum(residuals)
                    #postscript(file="../../results/2014_sep_output_gap/diff_mean_Gap2.eps",horizontal=FALSE,width=10,
png(file="../../results/2014_sep_output_gap/diff_mean_Gap2.png",width=800,height=400)
plot(diff)
dev.off()


gap_cbo <- qGDP-qPotGDP

data <- merge(qeRm,lag(gap_cbo,-2),all=FALSE)
head(data)
EqtyPrem <- data$qeRm
names(EqtyPrem)<- "eRm"
GapCBO <- data[,2]
names(GapCBO)<- "GapCBO"

regrCBO <- lm(EqtyPrem ~ GapCBO)
summary(regrCBO)

demeaned_eRm <- (EqtyPrem - mean(EqtyPrem))^2
residuals <- regrCBO$residuals^2

head(demeaned_eRm)
head(residuals)
```

```
diff <- cumsum(demeaned_eRm)-cumsum(residuals)
#postscript(file="../../results/2014_sep_output_gap/diff_mean_GapCBO.eps",horizontal=FALSE,width=10,height=5)
png(file="../../results/2014_sep_output_gap/diff_mean_GapCBO.png",width=800,height=400)
plot(diff)
dev.off()

stargazer(regr1,regr2,regrCBO,out="../../results/2014_sep_output_gap/in_sample_predictability_gap.tex",
          float=FALSE,
          omit.stat=c("f","rsq","ser"))

postscript(file="../../results/2014_sep_output_gap/ts_Gap1.eps",horizontal=FALSE,width=10,height=5)
png(file="../../results/2014_sep_output_gap/ts_Gap1.png",width=800,height=400)
plot(Gap1)
dev.off()

                                        #postscript(file="../../results/2014_sep_output_gap/ts_Gap2.eps",horizontal=FALSE,width=10,height=
png(file="../../results/2014_sep_output_gap/ts_Gap2.png",width=800,height=400)
plot(Gap2)
dev.off()

                                        #postscript(file="../../results/2014_sep_output_gap/ts_GapCBO.eps",horizontal=FALSE,width=10,heig
png(file="../../results/2014_sep_output_gap/ts_GapCBO.png",width=800,height=400)
plot(GapCBO)
dev.off()
```

5 Literature

Some central references

Biases in estimators of predictive regressions

Stambaugh (1999), Nelson and Kim (1993), Ferson, Sarkissian, and Simin (2003), Lewellen (2004)

Usefulness of predictability for asset pricing Kandel and Stambaugh (1996), Stambaugh (1999)

Hodrick (1992) correct standard errors for long term predictability, also Ang and Bekaert (2007)

Bayesian perspective Cremers (2002)

Summary status 2008: Goyal and Welch (2008) - no predictability

Lettau and van Nieuwerburgh (2008) argues against Goyal and Welch (2008), show that removing structural breaks will restore predictability

Another critical piece to the Goyal Welch analysis is Campbell and Thompson (2008), which argue that if one does some sensible restrictions on predictions, such as imposing that the equity premium is nonnegative, regains some predictaility

Cochrane (2008) (rfs) question power of Goyal and Welch (2008), argues that there must be predictability from dividend price ratio movement.

Chen (2009) returns predictability concentrated in postwar data

Henkel et al. (2011): Predictability concentrates in business cycle contration periods

Recent survey: Rapach and Zhou (2013)

Comparing model based expectations (like those investigated in Goyal and Welch (2008)), to surveys of investor expectations. In particular find negative correlations between model-based expectations to investor forecast. Argue that the investor forecasts are to extrapolative. Show that investors trade on their expectations. Question: Who is on the other side?

References

Andrew Ang and Gert Bekaert. Stock return predictability. Is it there? *Review of Financial Studies*, 20(3):423–458, 2007.

John Y Campbell and Samuel B Thompson. Predicting excess stock returns out of sample: Can anything beat the historical average. *Review of Financial Studies*, 21(4):1509–1532, 2008.

Long Chen. On the reversal of return and dividend growth predictability: A tale of two periods. *Journal of Financial Economics*, 92(1):128–151, 2009.

John Cochrane. The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21(4): 1533–1575, 2008.

Ilan Cooper and Richard Priestley. Time-varying risk premiums and the output gap. *Review of Financial Studies*, 2008.

K J Martin Cremers. Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies*, 15(4):1223–1249, 2002.

Wayne E Ferson, Sergei Sarkissian, and Timothy T Simin. Spurios regressions in financial economics? *Journal of Finance*, 58(4):1393–1414, 2003.

Amit Goyal and Ivo Welch. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508, 2008.

Sam James Henkel, J Spencer Martin, and Federico Nardari. Time-varying short-horizon precitability. *Journal of Financial Economics*, 99(3):560–580, 2011.

Robert J Hodrick. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, 5(3):357–86, 1992.

Shmuel Kandel and Robert F Stambaugh. On the predictability of stock returns: An asset allocation perspective. *Journal of Finance*, 51(2):385–424, June 1996.

Martin Lettau and Stijn van Nieuwerburgh. Reconciling the return predictability evidence. *Review of Financial Studies*, 21(4):1607–1652, 2008.

Jonathan Lewellen. Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235, 2004.

Robert C Merton. On estimating the expected return on the market. *Journal of Financial Economics*, pages 323–362, 1980.

Charles R. Nelson and Myung J. Kim. Predictable stock returns: The role of small sample bias. *Journal of Finance*, 48 (2):641–661, June 1993.

David Rapach and Guofu Zhou. Forecasting stock returns. In *Handbook of Economic Forecasting*, volume 2A of *Handbooks in Economics*, chapter 6, pages 328–380. North-Holland, 2013.

Robert Stambaugh. Predictive regressions. *Journal of Financial Economics*, 54(3):375–421, December 1999.